

DOI [https://doi.org/10.15589/znp2021.1\(484\).12](https://doi.org/10.15589/znp2021.1(484).12)
УДК 004.8

SYSTEM OF IMAGE RECOGNITION OF SURROUNDING OBJECTS WITH NEURAL NETWORKS

СИСТЕМА РОЗПІЗНАВАННЯ ОБ'ЄКТІВ НАВКОЛИШНЬОГО СВІТУ НА ОСНОВІ НЕЙРОННОЇ МЕРЕЖІ

Kateryna O. Kirei
kirey.kea@gmail.com
ORCID: 0000-0002-9338-2380

К. О. Кірей,
канд. пед. наук, доцент

Petro Mohyla Black Sea National University, Mykolaiv
Чорноморський національний університет імені Петра Могили, м. Миколаїв

Abstract. The article considers one of the directions of automation of recognition of real world objects using the convolutional neural networks. The recognition of objects in an image and their classification is the basis of machine vision systems. However, real-world computer recognition systems do not always work well. The main problem is that in most cases it is not possible to properly identify the features on which recognition should be based. For tasks where such features can be identified, artificial recognition systems have become widespread and widely used. For example, recognition of vehicle license plates, human faces, etc. However, sufficient adequacy of selection and recognition of various objects of the real world in the time close to the real one is still not achieved, which does not allow to achieve the necessary indicators in real tasks.

Purpose of research is the development of an efficient neural network algorithm capable of quickly and accurately recognizing real-world objects containing images from a predetermined list of objects.

Method. based on the analysis of algorithms and neural network architecture, the most appropriate type of neural network was chosen; the algorithm of object recognition in the image was optimized; the most suitable configuration of hyperparameterizations of neural networks was empirically chosen, directly influences the speed and performance of the neural network in recognition.

Results. Several models have been developed for neural networks, such as precise, precise2, and fast, with specific hyperparameter configurations. Then the neural network was trained on the basis of these configurations with the COCO dataset and the main functions of the created system were tested.

Scientific novelty is the development of an advanced algorithm for the recognition of objects and their borders on the image.

Practical importance. A system has been developed that can quickly and accurately recognize and mark objects in an image. The system can also work with any data set and is very flexible in setting.

Key words: computer vision; convolutional neural network; region proposals; regions; hyperparameters.

Анотація. У статті розглядається один із напрямів автоматизації розпізнавання образів з використанням згорткових нейронних мереж. Розпізнавання образів на зображенні та їх класифікація є основою систем машинного зору. Однак комп'ютерні системи розпізнавання образів реального світу не завжди працюють добре. Головна проблема полягає у тому, що в більшості випадків неможливо адекватно визначити ознаки, на основі яких слід здійснювати розпізнавання. Для задач, де такі ознаки вдається виділити, штучні системи розпізнавання набули значного поширення і широко використовуються. Наприклад, розпізнавання номерних знаків транспортних засобів, обличчя людини тощо. Проте все ще не досягнута достатня адекватність виділення і розпізнавання різноманітних об'єктів реального світу у часі, наближеному до реального, що не дозволяє досягти необхідних показників в реальних задачах.

Метою дослідження є розробка алгоритму ефективної нейронної мережі, здатної швидко та точно розпізнавати об'єкти реального світу, що містить зображення, з заздалегідь визначеного переліку об'єктів.

Методика. На основі аналізу алгоритмів та архітектур нейронних мереж було обрано найбільш придатний тип нейронної мережі; оптимізовано алгоритм розпізнавання об'єктів на зображенні; емпіричним шляхом обрано найбільш прийнятну конфігурацію гіперпараметрів згорткових нейронних мереж, що напряму впливає на швидкість та якість роботи нейронної мережі у розпізнаванні.

Результати. Було розроблено декілька моделей згорткових нейронних мереж: precise, precise2 та fast, що мають певну конфігурацію гіперпараметрів. Після чого було навчено нейронну мережу на основі цих конфігурацій із датасетом СОСО та проведено тестування основних функцій створеної системи.

Наукова новизна полягає у розробці вдосконаленого алгоритму розпізнавання об'єктів та їх меж на зображенні.

Практична значимість. Розроблено систему, здатну швидко та точно розпізнавати та позначати об'єкти на зображенні. Також система може працювати з будь-якими наборами даних та є дуже гнучкою у налаштуванні.

Ключові слова: комп'ютерний зір; згорткова нейронна мережа; пропозиції регіонів; регіони; гіперпараметри.

ПОСТАНОВКА ЗАДАЧІ

Потужність сучасних комп'ютерів та великий стрибок у сфері розробки штучного інтелекту за останні роки роблять можливим все більший розвиток такої технології, як машинний зір. Розпізнавання образів на зображенні та їх класифікація є основою систем машинного зору. Сьогодні застосування цієї технології може бути найрізноманітнішим: від розпізнавання машинами об'єктів на конвеєрі у промисловому виробництві до виявлення ракових клітин у біопсії тканини. Традиційно задачі розпізнавання образів включають у коло задач штучного інтелекту [1]. Для їх вирішення використовується величезна кількість різних методів: виділення країв, колірне сегментування зображення, застосування штучних нейронних мереж, використання особливих точок, методи порівняння з еталоном, методи генерації ознак. Взагалі, розпізнавання образів – це віднесення вихідних даних до певного класу за допомогою виділення істотних ознак, що характеризують ці дані, із загальної маси несуттєвих даних. Проблеми розпізнавання легко вирішуються людським мозком і робиться це, зазвичай, підсвідомо. Однак комп'ютерні системи розпізнавання об'єктів реального світу не завжди працюють добре. Головна проблема полягає у тому, що у більшості випадків неможливо адекватно визначити ознаки, на основі яких слід здійснювати розпізнавання. Для задач, де такі ознаки вдається виділити, штучні системи розпізнавання набули значного поширення і широко використовуються. Наприклад, розпізнавання номерних знаків транспортних засобів, обличчя людини тощо. Проте що все ще не досягнута достатня адекватність виділення і розпізнавання різноманітних об'єктів реального світу у часі, наближеному до реального, що не дозволяє досягти необхідних показників в реальних задачах.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Проблемами комп'ютерного зору наука активно переймається з кінця 1970-х років. Все це знайшло відображення у чисельних працях вітчизняних та закордонних науковців, таких як Р. Гиршик, Т. Даррелл, С. Диввала, І. Добротвор, Дж. Донахью, В. Кутковецький, Дж. Малик, Дж. Редмон, П. Стухляк, Р. Фархаді, М. Фісун тощо. Так, наприклад, для оптичного розпізнавання образів можна застосувати метод перебору вигляду об'єкта під різними кутами, масштабами,

зсувами. Для букв потрібно перебирати шрифт, властивості шрифту. Другий підхід – знайти контур об'єкта й досліджувати його властивості (зв'язність, наявність кутів тощо). Ще один підхід – використовувати нейронні мережі [2; 3; 4], зокрема згорткові нейронні мережі (ЗНМ). Цей підхід вимагає або великої кількості прикладів задачі розпізнавання (із правильними відповідями), або спеціальної структури нейронної мережі, що враховує специфіку певної задачі. Також заслуговує на увагу метод, заснований на моделі мозку, яку описав Ф. Розенблатт [5]. Завдання цієї моделі полягає в тому, щоб показати, як у деякій фізичній системі, структура й функціональні властивості якої відомі, можуть виникати психологічні явища. Ф. Розенблатт описав найпростіші експерименти з розрізнення. Дані експерименти цілком стосуються методів розпізнавання образів, але відрізняються тим, що алгоритм розв'язання не детермінований. Нині у світі набули широкого поширення ряд методів та алгоритмів, що застосовуються для вирішення задач розпізнавання образів. Так у роботах [6; 7; 8] описано алгоритми, що засновані на підході Region Proposals, тобто виділені регіонів, на яких імовірно є об'єкти припущень. У роботах [9; 10] описано алгоритми, що засновані на обробці зображення в цілому.

МЕТА ДОСЛІДЖЕННЯ

Метою дослідження є розробка алгоритму ефективною нейронної мережі, здатної швидко та точно розпізнавати об'єкти реального світу, що містить зображення, з заздалегідь визначеного переліку об'єктів.

МЕТОДИ, ОБ'ЄКТ ТА ПРЕДМЕТ ДОСЛІДЖЕННЯ

Для досягнення мети на основі аналізу алгоритмів та архітектур нейронних мереж було обрано найбільш придатний тип нейронної мережі. Емпіричним шляхом обрано найбільш прийнятну конфігурацію моделі, що напряму впливає на швидкість та якість роботи нейронної мережі у розпізнаванні. Оптимізовано алгоритм розпізнавання об'єктів на зображенні.

Об'єктом дослідження є автоматизоване розпізнавання об'єктів реального світу.

Предметом дослідження є розпізнавання об'єктів на основі ЗНМ.

ОСНОВНИЙ МАТЕРІАЛ

Під образом, що має бути розпізнаним, будемо розуміти об'єкт, процес або явище реального

та абстрактного світу, який розпізнається за даними (ознаками), що збираються та оброблюються індивідуально і у сукупності [11, с. 9]. Метод розпізнавання образів з використанням нейронних мереж має суттєві переваги перед іншими методами розпізнавання. Методи розпізнавання об'єктів реального світу за формою контурів сьогодні майже не використовуються з двох основних причин: об'єкт в реальних умовах може не мати чіткого контуру, зливатися з навколишнім фоном або перетинатися з іншими, нецільовими об'єктами, що помітно ускладнює не тільки розпізнавання об'єкта, але і точну побудову його форми. Морфологічні перетворення також більшою мірою підходять для роботи з символами, ніж зі складними образами реального світу. Сьогодні для завдання розпізнавання об'єктів реального світу найкращим чином підходять згорткові нейронні мережі (ЗНМ, англ. Convolutional Neural Network, CNN, ConvNet) в машинному навчанні – це клас глибоких штучних нейронних мереж прямого поширення, який успішно застосовувався до аналізу візуальних зображень [12]. ЗНМ призначені для розпізнавання візуальних патернів безпосередньо з піксельних зображень з мінімальною попередньою обробкою, що значно прискорює процес розпізнавання [13].

Більшість систем розпізнавання на основі ЗНМ застосовують модель до зображення в декількох його місцях та масштабах. Регіони зображення з високою оцінкою вважаються розпізнаними. Прикладами таких систем можуть слугувати системи R-CNN (Region-based Convolutional Network), Fast R-CNN та Faster R-CNN [8; 14]. Алгоритм Faster R-CNN є логічним продовженням алгоритму Fast R-CNN [6], в якому вирішені проблеми довгого і не дуже точного пошуку вікон-кандидатів, Faster R-CNN вважається одним з найбільш точних алгоритмів виявлення об'єктів на зображенні за допомогою нейромереж, але для досягнення високої точності доводиться жертвувати швидкістю обробки, що є неприйнятним для використання в режимі, наближеному до реального. Задля досягнення великої швидкості розпізнавання розроблена нами система базується на підході, запропонованому у роботі [9]. Ідея полягає у застосованні єдиної нейронної мережі до цілого зображення одразу. Ця нейронна мережа розбиває зображення на умовні регіони і передбачає межі розпізнаного об'єкта та ймовірності для кожного регіону. Ці межі є зваженими відносно передбачених ймовірностей. Тобто підґрунтям пропонованого алгоритму розпізнавання є те, що нейронна мережа «дивиться» на зображення лише один раз, після чого виконує всі обчислення. Система отримує вхідне зображення, змінює його розміри до таких, які зазначено у файлі конфігурації моделі, та розбиває зображення на клітини (рис. 1).



Рис. 1. Розбиття зображення на клітини



Рис. 2. Передбачення меж та ймовірностей для двох клітин



Рис. 3. Передбачення меж та ймовірностей для всіх клітин



Рис. 4. Передбачення ймовірностей класів

Всі клітини передаються в ЗНМ, де для кожної з них передбачаються межі об'єктів та ймовірності (рис. 2-3). Також для кожної клітини передбачається ймовірність належності до певного класу (рис. 4).

Останнім етапом алгоритму є комбінування визначених меж із визначеними класами та відкидання результатів із низькими ймовірностями (рис. 5).

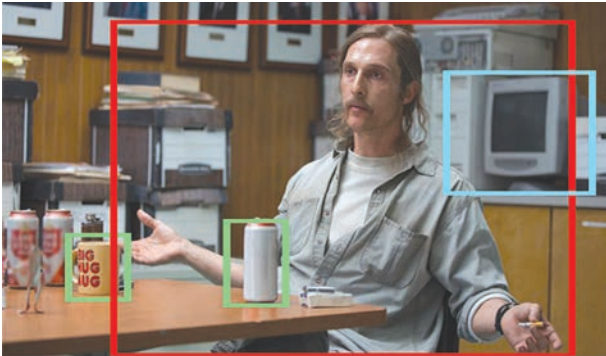


Рис. 5. Результат передбачень системи

Найголовніша функція пропонованої системи – власне розпізнавання об'єктів на зображенні. Для того, щоб запустити розпізнавання (detect), необхідно у командному рядку виконати таку команду: `./observer detect <файл даних> <файл конфігурації> <файл ваг> <файл зображення до розпізнавання> [-thresh (0.0; 1.0)]`. Необов'язковий параметр `-thresh` вказує на бажаний поріг ідентифікації об'єкта як розпізнаного, за замовченням цей параметр має значення 0,3. В результаті виконання такої команди у кореновому каталозі програми створюється файл `predictions.jpg`, на якому відмічено знайдені об'єкти (рис. 6).

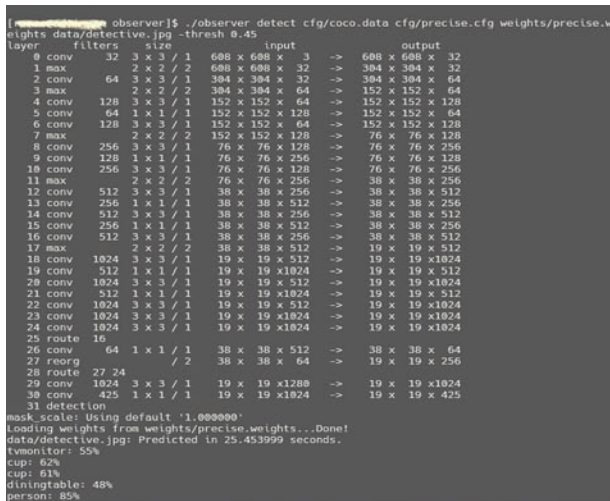


Рис. 6. Результат виконання стандартної команди розпізнавання із моделлю precise

Важливим етапом є вибір найбільш прийнятної конфігурації моделі, що напряму впливає на швидкість та якість роботи нейронної мережі у розпізнаванні. Є декілька загальноновизначених конфігурацій,

наприклад, таких як AlexNet, але можна розробити і свою власну. На рис. 8 наведено приклад файлу конфігурації.



Рис. 7. Візуалізація передбачень моделі precise

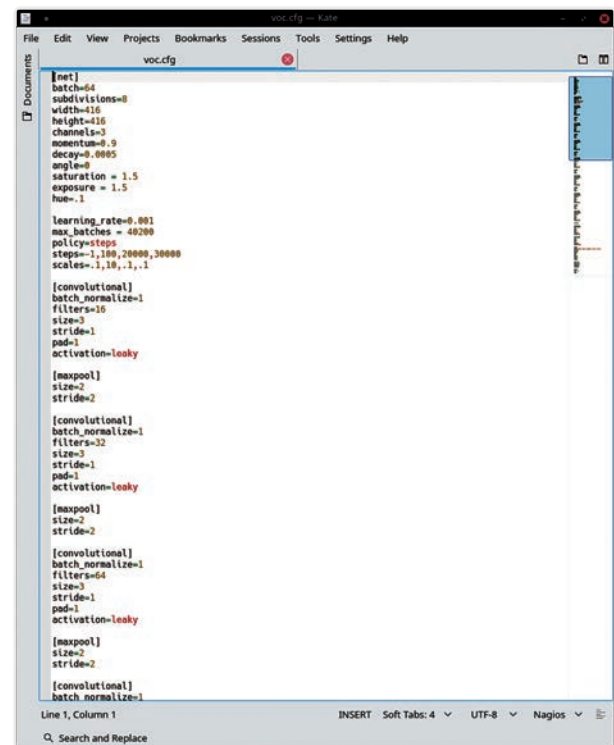


Рис. 8. Файл конфігурації моделі

Отже, пропонована система має декілька переваг, порівняно із системами, що базуються на класифікаторах. Вона «дивиться» одразу на ціле зображення, що надає їй змогу урахувати глобальний контекст у зображенні під час передбачення. Вона також виконує передбачення лише за один цикл обчислень мережі, що робить її дуже швидкою.

Було розроблено декілька моделей ЗНМ з різною конфігурацією гіперпараметрів: `precise`, `precise2` та `fast`. Всі моделі створювалися емпірично з урахуванням особливостей ЗНМ. Після чого було навчено нейронну мережу на основі цих моделей із датасетом COCO (Common Objects in Context) [15; 16]. Цей датасет

включає в себе більше 200 тис. підписаних зображень, які охоплюють 80 категорій об'єктів.

ОБГОВОРЕННЯ РЕЗУЛЬТАТІВ

Приклад роботи моделі *precise* наведено вище. Інша модель, *precise2*, є більш оптимізованою версією моделі *precise*. Її швидкість роботи вища приблизно в 2 рази, а точність розпізнавання лише трохи гірша. На рис. 8 наведено результати роботи моделі *precise2*.

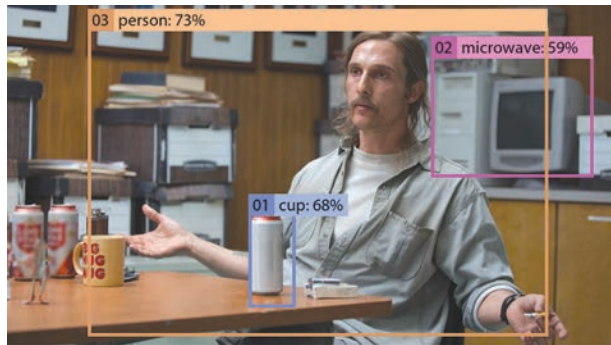


Рис. 8. Візуалізація передбачень моделі *precise2*

Третя модель – *fast* – відпрацьовує на порядок швидше навіть за модель *precise2*, але при цьому жертвує точністю, втім, в деяких випадках вона може застосовуватись дуже ефективно, наприклад, в зображеннях із невеликою кількістю об'єктів. Приклади роботи моделі *fast* наведено на рис. 9.

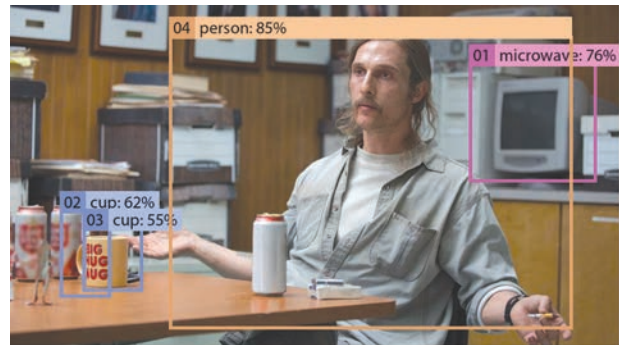


Рис. 9. Візуалізація передбачень моделі *fast*

ВИСНОВКИ

Запропонований підхід щодо розпізнавання об'єктів реального світу дає змогу не тільки використовувати завчасно розроблені та натреновані моделі, можна також створювати і свої моделі мережі, застосовувати довільні датасети та самостійно навчати нейронну мережу. Користувач може одночасно мати скільки завгодно готових до використання натренованих моделей, і використовувати їх без необхідності кожного разу заново навчати нейронну мережу. Адже процес навчання є дуже довгою процедурою і перш за все залежить від розміру датасету. Запропонована система розпізнавання об'єктів може бути використана там, де є потреба здійснювати контроль за певними об'єктами. Наприклад, контроль складських або торгових приміщень – автоматизація контролю того, що вносять або виносять, наявності заборонених або підозрілих об'єктів тощо.

REFERENCES

- [1] Wikipedia. (2020) Teoriia rozpoznavannia obraziv [Theory of Pattern Recognition]. In *Wikipedia – The Free Encyclopedia*. Retrieved from https://uk.wikipedia.org/wiki/Теорія_розпізнавання_образів. [in Ukrainian]
- [2] Tihanov, O., Kravets, I., Fisun, M. (2017). Systema invariantnoho rozpoznavannia ob'ektiv na zaliznychnomu transporti na osnovi neironnykh merezh [The System Of Object's Invariant Recognition For Railway Transport Based On Neural Networks]. *Inzhenerni nauky*, vol. 1 (16), pp. 226-232. [in Ukrainian]
- [3] Dobrotvor, I.H., Stukhliak, P.D., Myktyshyn, A.H. & Mytnyk, M.M. (2018). *Analiz system rozpoznavannia obraziv struktury kompozytiv : monohrafiia* [Analysis of composite structure pattern recognition systems]. Ternopil : Ternopilskyi natsionalnyi tekhnichnyi universytet imeni Ivana Puliuiia. Retrieved from <http://elartu.tntu.edu.ua/handle/lib/25615>. [in Ukrainian]
- [4] Basiuk, T., Pushko, Ya. (2015). Analiz ta klasyfikatsiia osnovnykh metodiv rozpoznavannia obraziv na ploshchyni proektsii [Analysis and classification of the main methods of image recognition on the projection plane]. Lviv Polytechnic National University Institutional Repository, <http://ena.lp.edu.ua:8080/bitstream/ntb/31327/1/40-291-298.pdf>. [in Ukrainian]
- [5] Rosenblatt, F. (1962). Principles Of Neurodynamics: Perceptrons And The Theory Of Brain Mechanisms. Spartan Books.
- [6] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. CVPR, v5. Retrieved from <https://arxiv.org/abs/1311.2524>.
- [7] Girshick, R., Donahue, J., Darrell, T. & Malik, J. (2016, January). Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans Pattern Anal Mach Intell*, 38(1):142-58. doi: 10.1109/TPAMI.2015.2437384.
- [8] Girshick, R. (2015, December). Fast R-CNN. *IEEE International Conference on Computer Vision (ICCV)*, pp. 1440-1448. doi: 10.1109/ICCV.2015.169.
- [9] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016, May). You Only Look Once: Unified, Real-Time Object Detection. *arXiv*. Retrieved from <https://arxiv.org/pdf/1506.02640>.
- [10] Redmon, J., Farhadi, A. (2016, Desember). YOLO9000: Better, Faster, Stronger. *arXiv*. Retrieved from <https://arxiv.org/pdf/1612.08242>.
- [11] Kutkovetskyi, V.Ia. (2017). *Rozpoznavannia obraziv : navch. posib.* [Pattern Recognition]. Mykolaiv : ChNU im. Petra Mohyly, 2017. [in Ukrainian]
- [12] Wikipedia. (2020). Zghortkova neironna mreza [Convolutional neural network]. *Wikipedia – The Free Encyclopedia*. Retrieved from https://uk.wikipedia.org/wiki/Згорткова_нейронна_мережа. [in Ukrainian]

- [13] LeCun, Y., Bottou, L., Bengio, Y. & Haffner P. (1998, November). GradientBased Learning Applied to Document Recognition. *Proc. of IEEE*, vol. 86 (11), pp. 2278-2324. doi: 10.1109/5.726791.
- [14] Ren, S., He, K., Girshick, R. & Sun, J. (2017, June). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39 (6), pp. 1137-1149. doi:10.1109/TPAMI.2016.2577031.
- [15] COCO – Common Objects in Context. (2021). *COCO 2020*. COCO – Common Objects in Context. <https://cocodataset.org/#home>.
- [16] Lin, T.-Yi, Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., ... Dollár, P. (2015, February). Microsoft COCO: Common Objects in Context. *arXiv*. Retrieved from <https://arxiv.org/abs/1405.0312>.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] Вікіпедія. (2020). Теорія розпізнавання образів. *Матеріал з Вікіпедії — вільної енциклопедії*. URL : https://uk.wikipedia.org/wiki/Теорія_розпізнавання_образів.
- [2] Тіганов О.С., Кравець І.О., Фісун М.Т. Система інваріантного розпізнавання об'єктів на залізничному транспорті на основі нейронних мереж. *Інженерні науки*. 2017. № 1(16). С. 226–232.
- [3] Добротвор, І.Г., Стухляк, П.Д., Микитишин, А.Г. та Митник, М.М. (2018). *Аналіз систем розпізнавання образів структури композитів : монографія*. Тернопіль : Тернопільський національний технічний університет імені Івана Пулюя. URL : <http://elartu.tntu.edu.ua/handle/lib/25615>.
- [4] Басюк, Т., Пушко, Я. (2015). Аналіз та класифікація основних методів розпізнавання образів на площині проєкції. *Lviv Polytechnic National University Institutional Repository*. URL : <http://ena.lp.edu.ua:8080/bitstream/ntb/31327/1/40-291-298.pdf>.
- [5] Rosenblatt, F. (1962). *Principles Of Neurodynamics: Perceptrons And The Theory Of Brain Mechanisms*. Spartan Books.
- [6] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *CVPR*, v5. URL : <https://arxiv.org/abs/1311.2524>.
- [7] Girshick, R., Donahue, J., Darrell, T. & Malik, J. (2016, January). Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans Pattern Anal Mach Intell*, 38(1):142-58. doi: 10.1109/TPAMI.2015.2437384.
- [8] Girshick, R. (2015, December). Fast R-CNN. *IEEE International Conference on Computer Vision (ICCV)*, pp. 1440-1448. doi: 10.1109/ICCV.2015.169.
- [9] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016, May). You Only Look Once: Unified, Real-Time Object Detection. *arXiv*. Retrieved from <https://arxiv.org/pdf/1506.02640>.
- [10] Redmon, J., Farhadi, A. (2016, Desember). YOLO9000: Better, Faster, Stronger. *arXiv*. Retrieved from <https://arxiv.org/pdf/1612.08242>.
- [11] Кутковецький, В. Я. (2017). *Розпізнавання образів : навч. посіб.* Миколаїв: ЧНУ ім. Петра Могили.
- [12] Вікіпедія. (2020). Згортовка нейронна мережа. *Матеріал з Вікіпедії — вільної енциклопедії*. URL : https://uk.wikipedia.org/wiki/Згортовка_нейронна_мережа.
- [13] LeCun, Y., Bottou, L., Bengio, Y. & Haffner P. (1998, November). GradientBased Learning Applied to Document Recognition. *Proc. of IEEE*, vol. 86 (11), pp. 2278-2324. doi: 10.1109/5.726791.
- [14] Ren, S., He, K., Girshick, R. & Sun, J. (2017, June). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39 (6), pp. 1137-1149. doi:10.1109/TPAMI.2016.2577031.
- [15] COCO – Common Objects in Context. (2021). *COCO 2020*. COCO – Common Objects in Context. URL : <https://cocodataset.org/#home>.
- [16] Lin, T.-Yi, Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., ... Dollár, P. (2015, February). Microsoft COCO: Common Objects in Context. *arXiv*. URL : <https://arxiv.org/abs/1405.0312>.

© Кірей К. О.

Дата надходження статті до редакції: 10.03.2021

Дата затвердження статті до друку: 24.03.2021